

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-084255

(43)Date of publication of application : 30.03.2001

(51)Int.Cl.

G06F 17/30

(21)Application number : 11-257665

(71)Applicant : FUJI XEROX CO LTD

(22)Date of filing : 10.09.1999

(72)Inventor : UMEKI HIROSHI

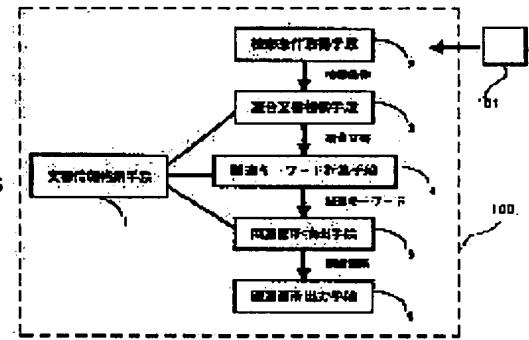
TATENO SHOICHI

(54) DEVICE AND METHOD FOR RETRIEVING DOCUMENT

(57)Abstract:

PROBLEM TO BE SOLVED: To extract a relevant position in watching with a retrieval intention from a retrieval document.

SOLUTION: A retrieval condition acquiring means 2 receives a logical operation expression for a keyword from a user as a retrieval condition. A compatible document retrieving means 3 finds a list of document IDs corresponding to each inputted keyword from a ward index of a document information storing means 1, performs a designated logical operation and calculates a compatible document ID. A relevant keyword calculating means 4 calculates the degree of association about each compatible document with a keyword extracted from the compatible document retrieved by the means 3 as an associated keyword. A relevant position extracting means 5 accumulates the degrees of association of a relevant keyword for each statement of the compatible document and successively extracts sentences from a sentence, having a large accumulated value until being equal to or larger than a prescribed length, while keeping the sequences which appears in the document. A relevant place outputting means 6 displays a statement found from the means 5 as retrieval results to a user.



(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2001-84255

(P2001-84255A)

(43)公開日 平成13年3月30日 (2001.3.30)

(51)IntCL
G 06 F 17/30

識別記号

F I
G 06 F 16/403
16/40
15/401

3 8 0 D

テ-マコト(参考)
5 B 0 7 5
3 7 0 A
3 2 0 A

審査請求 未請求 請求項の数17 OL (全 20 頁)

(21)出願番号 特願平11-257665

(22)出願日 平成11年9月10日 (1999.9.10)

(71)出願人 000005496

富士ゼロックス株式会社
東京都港区赤坂二丁目17番22号

(72)発明者 梅基 宏

神奈川県足柄上郡中井町境430 グリーン
テクなかい 富士ゼロックス株式会社内

(72)発明者 館野 崑一

神奈川県足柄上郡中井町境430 グリーン
テクなかい 富士ゼロックス株式会社内

(74)代理人 100086531

弁理士 淳田 俊夫

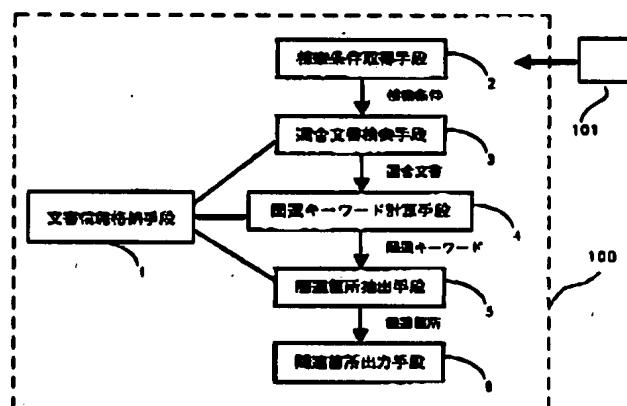
Fターム(参考) 5B075 ND03 NK02 PP02 PP22 PQ02
PQ32 PQ75 PR08 UU06

(54)【発明の名称】 文書検索装置および方法

(57)【要約】

【課題】 検索文書から検索意図に合致した関連箇所を抽出する。

【解決手段】 検索条件取得手段2は、ユーザから検索条件としてキーワードの論理演算式を受け取る。適合文書検索手段3は、入力された各キーワードに対応する文書IDのリストを文書情報格納手段1の単語索引から求め、指定の論理演算を施し、適合する文書IDを求める。関連キーワード計算手段4は、適合文書検索手段3で検索された適合文書から抽出されたキーワードを関連キーワードとして、各々についてその関連度を計算する。関連箇所抽出手段5は、適合文書の各文について関連キーワードの関連度を累積し、その累積値の大きい文章から順に所定の長さ以上になるまで、文書に現れる順番を保ったまま文章を抽出していく。関連箇所出力手段6は、関連箇所抽出手段5から求まる文を、検索結果としてユーザに対して表示する。



(2)

特開2001-84255

【特許請求の範囲】

【請求項1】 入力された検索条件に適合する文書を検索する文書検索装置において、複数の文書を、各文書から抽出されたキーワードと対応づけて格納する文書情報格納手段と、検索条件を受け取る検索条件取得手段と、前記文書情報格納手段に格納されている文書を対象として、前記検索条件取得手段が受け取った検索条件に適合する適合文書を検索する適合文書検索手段と、前記適合文書検索手段で検索された適合文書に対応して前記文書情報格納手段に格納されているキーワードを関連キーワードとして求め、前記関連キーワードの各々について前記適合文書の中で該関連キーワードを含んでいる文書の数と、前記文書情報格納手段に格納されている文書の中で該関連キーワードを含んでいる文書の数とを変数とする計算式に基づいて、前記検索条件取得手段が受け取った検索条件と該関連キーワードとの間の関連度を算出する関連キーワード計算手段と、

前記関連キーワード計算手段から得られる関連キーワードとその関連度に基づいて、前記適合文書の内容から関連箇所を抽出する関連箇所抽出手段と、前記関連箇所抽出手段から得られる関連箇所を出力する関連箇所出力手段とを有することを特徴とする文書検索装置。

【請求項2】 前記文書情報格納手段に格納されている文書を対象として、前記関連キーワード計算手段から得られる関連キーワードとその関連度に基づいて、前記検索条件取得手段が受け取った検索条件に関連する関連文書を検索する関連文書検索手段を有し、

前記関連箇所抽出手段は、前記関連キーワード計算手段から得られる関連キーワードとその関連度に基づいて、前記関連文書検索手段から得られる関連文書の内容から関連箇所を抽出する請求項1記載の文書検索装置。

【請求項3】 入力された検索条件に関連する文書を検索する文書検索装置において、複数の文書を、各文書から抽出されたキーワードと対応づけて格納する文書情報格納手段と、検索条件を受け取る検索条件取得手段と、前記文書情報格納手段に格納されているキーワードの中から、関連度を判断する対象とすべき関連キーワードを特定し、前記文書情報格納手段に格納されている文書の中で関連キーワードを含んでいる文書の数を変数とする計算式に基づいて、前記検索条件取得手段が受け取った検索条件と関連キーワードとの間の関連度を算出する関連キーワード計算手段と、

前記文書情報格納手段に格納されている文書を対象として、前記関連キーワード計算手段から得られる関連キーワードとその関連度に基づいて、前記検索条件取得手段が受け取った検索条件に関連する関連文書を検索する関連文書検索手段と、

前記関連キーワード計算手段から得られる関連キーワードとその関連度に基づいて、前記関連文書検索手段から得られる関連文書の内容から関連箇所を抽出する関連箇所抽出手段と、

前記関連箇所抽出手段から得られる関連箇所を出力する関連箇所出力手段とを有することを特徴とする文書検索装置。

【請求項4】 前記検索条件取得手段は、前記文書情報格納手段に格納されている一つあるいは複数の文書を入力として受け取り、

前記関連キーワード計算手段は、前記検索条件取得手段が受け取った文書に含まれることばを関連キーワードとし、前記検索条件取得手段が受け取った検索条件と関連キーワードとの間の関連度を、前記文書の中で該関連キーワードを含む文書の数と、前記文書情報格納手段に格納されている文書の中の該関連キーワードを含む文書の数とを変数とする計算式に基づいて算出する請求項3記載の文書検索装置。

【請求項5】 前記検索条件取得手段は、自然文を入力として受け取り、

前記関連キーワード計算手段は、前記文書情報格納手段に格納されているキーワードの中から、前記検索条件取得手段が受け取った自然文に含まれることばを関連キーワードとする請求項3記載の文書検索装置。

【請求項6】 前記関連文書検索手段から検索された関連文書を、前記関連箇所出力手段から出力される関連箇所と対応づけて出力する文書出力手段を有する請求項1、2、3、4または5記載の文書検索装置。

【請求項7】 前記関連箇所出力手段および前記文書出力手段は、

それらの出力に含まれる、前記関連キーワード計算手段から得られる関連キーワードの出力の形態を、該関連キーワードの関連度に応じて変える請求項1、2、3、4、5または6記載の文書検索装置。

【請求項8】 複数の文書を、各文書から抽出されたキーワードと対応づけて格納する文書情報格納手段と、検索条件を受け取る検索条件取得手段と、

前記文書情報格納手段に格納されている文書を対象として、前記検索条件取得手段が受け取った検索条件に適合する適合文書を検索する適合文書検索手段と、

前記適合文書検索手段で検索された適合文書に対応して前記文書情報格納手段に格納されているキーワードを関連キーワードとして求め、前記関連キーワードの各々について前記適合文書の中で該キーワードを含んでいる文書の数と、前記文書情報格納手段に格納されている文書の中で該キーワードを含んでいる文書の数とを変数とする計算式に基づいて、前記検索条件取得手段が受け取った検索条件と該キーワードとの間の関連度を算出し、関連キーワードとその関連度を求める関連キーワード計算手段と、

(3)

特開2001-84255

前記関連キーワード計算手段から得られる関連キーワードとその関連度に基づいて、前記文書情報格納手段に格納されている文書を対象として、前記検索条件に関連する関連文書を検索する関連文書検索手段とを有することを特徴とする文書検索装置。

【請求項9】 複数の文章からなる入力文書の各文章を所定の評価基準によりスコアを割り当てる手段と、上記スコアに基づいて順次に文章を抽出する手段と、抽出した文章の累積量が所定の分量基準を上回ったときに抽出を打ち切る手段と、抽出された文章を出力文書にまとめて出力する手段とを有することを特徴とする文書処理装置。

【請求項10】 上記分量基準は入力文書の量の固定割合とする請求項9記載の文書処理装置。

【請求項11】 前記打ち切る手段は、分量基準を上回った時点の文章まで抽出して出力文書に含ませる請求項9または10記載の文書処理装置。

【請求項12】 前記出力文書の文章は入力文書における各文章の出現順序で配置される請求項9、10または11記載の文書処理装置。

【請求項13】 前記出力文書の文章は前記評価基準のスコアに基づいて配置される請求項9、10、11または12記載の文書処理装置。

【請求項14】 前記打ち切る手段は、分量基準を上回る直前の文字まで抽出して出力文書に含ませる請求項9または10記載の文書処理装置。

【請求項15】 複数の文書を、各文書から抽出されたキーワードと対応づけて格納する文書情報格納手段と、検索条件を受け取る検索条件取得手段と、

前記文書情報格納手段に格納されている文書を対象として前記検索条件を利用した検索を実行する文書検索手段と、

前記検索により抽出された文書に対応して前記文書情報格納手段に格納されているキーワードを関連キーワードとして求め、前記関連キーワードの各々について前記適合文書の中で該キーワードを含んでいる文書の数と、前記文書情報格納手段に格納されている文書の中で該キーワードを含んでいる文書の数との少なくとも一方を変数とする計算式に基づいて、前記検索条件取得手段が受け取った検索条件と該キーワードとの間の関連度を算出する関連キーワード計算手段と、

前記関連キーワード計算手段から得られる関連キーワードとその関連度に基づいて、前記抽出された文書の内容から関連箇所を抽出する関連箇所抽出手段と、前記関連箇所抽出手段から得られる関連箇所を出力する関連箇所出力手段とを有することを特徴とする文書検索装置。

【請求項16】 複数の文書を、各文書から抽出されたキーワードと対応づけて格納する文書情報格納ステップと、検索条件を受け取る検索条件取得ステップと、

前記文書情報格納手段に格納されている文書を対象として前記検索条件を利用した検索を実行する文書検索ステップと、

前記検索により抽出された文書に対応して前記文書情報格納手段に格納されているキーワードを関連キーワードとして求め、前記関連キーワードの各々について前記適合文書の中で該キーワードを含んでいる文書の数と、前記文書情報格納手段に格納されている文書の中で該キーワードを含んでいる文書の数との少なくとも一方を変数とする計算式に基づいて、前記検索条件取得手段が受け取った検索条件と該キーワードとの間の関連度を算出する関連キーワード計算ステップと、

前記関連キーワード計算手段から得られる関連キーワードとその関連度に基づいて、前記抽出された文書の内容から関連箇所を抽出する関連箇所抽出ステップと、前記関連箇所抽出手段から得られる関連箇所を出力する関連箇所出力ステップとを有することを特徴とする文書検索装置。

【請求項17】 複数の文書を、各文書から抽出されたキーワードと対応づけて格納する文書情報格納ステップと、

検索条件を受け取る検索条件取得ステップと、前記文書情報格納手段に格納されている文書を対象として前記検索条件を利用した検索を実行する文書検索ステップと、

前記検索により抽出された文書に対応して前記文書情報格納手段に格納されているキーワードを関連キーワードとして求め、前記関連キーワードの各々について前記適合文書の中で該キーワードを含んでいる文書の数と、前記文書情報格納手段に格納されている文書の中で該キーワードを含んでいる文書の数との少なくとも一方を変数とする計算式に基づいて、前記検索条件取得手段が受け取った検索条件と該キーワードとの間の関連度を算出する関連キーワード計算ステップと、

前記関連キーワード計算手段から得られる関連キーワードとその関連度に基づいて、前記抽出された文書の内容から関連箇所を抽出する関連箇所抽出ステップと、前記関連箇所抽出手段から得られる関連箇所を出力する関連箇所出力ステップとをコンピュータシステムに実行させるためのコンピュータプログラムを記録した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は文書検索技術に関し、特に、検索された文書の内容から検索条件に関連している箇所を出力する文書検索技術に関する。

【0002】

【従来の技術】 各種的事项やキーワードなどを検索条件として文書を検索する従来の検索システムは、検索結果として一般に、検索結果の件数やタイトルリストなどを

(4)

特開2001-84255

表示する。検索された結果が検索意図に適合しているかどうかを判定するためには、検索された文書の各々の全文をユーザが参照して判断する必要があった。しかし、文書の全文にユーザの検索意図が必ずしも的確に表現されているとは限らず、また、多数の文書が検索された場合や文書の全文が長い場合は、全文をすべて読むために多大な時間が必要になってしまう。

【0003】近年、CD-ROMなどの大容量記憶メディアや、LANやインターネットなどのネットワークを通じて、大量の電子化文書が流通するようになり、これに伴い、大量の電子化文書を検索対象とする文書検索システムも一般的になった。しかし、そのような文書検索システムを使うと、大量の文書を検索されてしまうことがよくあるため、検索結果が適合しているかどうかを判定するために、ユーザに過剰な負担がかかってしまう場合が多い。

【0004】そこで、検索された文書の全文の一部だけを出力することで、そのような適合判定の負荷を軽くする方法が考えられる。

【0005】これまでに、テキストから自動的に要約を作成する様々な方法が提案されている。そのうちの一つは、テキストに頻出する名詞をキーワードと想定し、テキスト中の出現頻度に基づいて単語に重要度を与え、そのようにして求めた単語の重要度を元に、文に重要度を与え、重要な文を集めて要約とする方法である。また別の方法は、テキストの構造からテキスト中あるいは段落中で重要な箇所の位置をあらかじめ予測して、重要な文を抽出する方法である。

【0006】これらの方法では、同じテキストから常に同じ要約が作成されることになる。しかし、ユーザにとっては、同じテキストであっても異なる検索を行えば、検索意図が反映されて異なる要約を得られる方が望ましい。

【0007】一方、検索条件のキーワードを含んでいる検索結果の文書の近傍を抽出して表示する方法がある。この方法はKWIC (Keyword in Context) とよばれ、たとえばウェブの検索エンジンの表示などに広く使われている。しかし、検索条件に含まれるキーワードの数が少ない場合やキーワードの出現箇所が少ない場合、またはキーワードが的確に検索意図を表現していない場合、必ずしもそれらのキーワードの近傍だけにユーザの検索意図が表されているとは限らない。逆に、キーワードが多数の箇所に現れている場合、それらの箇所のどこがより重要なかを判定することが難しくなってしまう。

【0008】特開平10-207891号公報の「文書要約装置およびその方法」では、文書において重要な情報と、利用者が求める情報とを利用する文書要約方法が開示されている。この方法は、ユーザが興味を持った文書や重要と考えるキーワードなどをあらかじめ保存して

おき、ユーザの入力した検索条件やあらかじめ保存しておいたユーザの興味に関する情報などから、ユーザの興味を反映した要約を作成することを目的の一つとしている。しかし、この方法では、ユーザごとに興味に関する情報をあらかじめ入力しておき、適切に更新するなどの手間が必要になる。

【0009】

【発明が解決しようとする課題】以上述べたように、テキストの内容だけから文の重要度を求める従来の自動要約の技術では、ユーザの検索意図は考慮されない。

【0010】また、KWICでは、必ずしもキーワードの近傍だけにユーザの検索意図が表されているとは限らないことや、逆に、キーワードが多数の箇所に現れている場合、それらの箇所のどこがより重要なかを判定することは難しくなってしまう。

【0011】そして、特開平10-207891号公報の「文書要約装置およびその方法」のように、ユーザの興味に関する情報をあらかじめ入力しておく文書要約方法では、そのことによってユーザの興味は要約に反映される反面、ユーザごとに求める情報をあらかじめ入力しておくなどの手間が必要になってしまう。

【0012】本発明は、上記従来の事情に鑑みてなされたものであり、ユーザごとに興味のある情報をあらかじめ入力しておくなどの手間をかけずに、検索条件に関連する箇所を抽出して表示することができる文書検索装置を提供することを目的とする。

【0013】

【課題を解決するための手段】本発明は上述の目的を達成するために特許請求の範囲に記載のとおりの構成を採用している。

【0014】すなわち、本発明によれば、上述の目的を達成するために、入力された検索条件に適合する文書を検索する文書検索装置に：複数の文書を、各文書から抽出されたキーワードと対応づけて格納する文書情報格納手段と：検索条件を受け取る検索条件取得手段と；前記文書情報格納手段に格納されている文書を対象として、前記検索条件取得手段が受け取った検索条件に適合する適合文書を検索する適合文書検索手段と；前記適合文書検索手段で検索された適合文書に対応して前記文書情報格納手段に格納されているキーワードを関連キーワードとして求め、前記関連キーワードの各々について前記適合文書の中で該関連キーワードを含んでいる文書の数と、前記文書情報格納手段に格納されている文書の中で該関連キーワードを含んでいる文書の数とを変数とする計算式に基づいて、前記検索条件取得手段が受け取った検索条件と該関連キーワードとの間の関連度を算出する関連キーワード計算手段と；前記関連キーワード計算手段から得られる関連キーワードとその関連度に基づいて、前記適合文書の内容から関連箇所を抽出する関連箇所抽出手段と；前記関連箇所抽出手段から得られる関連

箇所を出力する関連箇所出力手段とを設けるようにしている。

【0015】この構成においては、関連キーワード、すなわち、適合文書（検索条件でヒットした文書）のいずれかのキーワードとなっていることばの各々について、適合文書中に出現する割合と全文書中で出現する割合とに基づいて、検索条件との関連度を求め、関連度の高いキーワードを含む文書部分を抽出するようにしているので、検索意図に合致した態様で文書部分を抽出することができる。

【0016】ここで、文書とは検索単位であり、1つの文章からなることもあるし、複数の文章からなることもある。

【0017】また、この構成において、文書検索装置に、さらに、前記文書情報格納手段に格納されている文書を対象として、前記関連キーワード計算手段から得られる関連キーワードとその関連度に基づいて、前記検索条件取得手段が受け取った検索条件に関連する関連文書を検索する関連文書検索手段を設け、前記関連箇所抽出手段が、前記関連キーワード計算手段から得られる関連キーワードとその関連度に基づいて、前記関連文書検索手段から得られる関連文書の内容から関連箇所を抽出するようにしてよい。

【0018】検索条件に関連し、その検索意図には合致するけれども、キーワードの割り振りに起因して検索条件には適合しない（ヒットしない）文書がある。関連キーワードと関連度とを用いれば、検索意図に合致する文書をより多く抽出できる。

【0019】また、本発明によれば、上述の目的を達成するために、入力された検索条件に関連する文書を検索する文書検索装置に：複数の文書を、各文書から抽出されたキーワードと対応づけて格納する文書情報格納手段と；検索条件を受け取る検索条件取得手段と；前記文書情報格納手段に格納されているキーワードの中から、関連度を判断する対象とすべき関連キーワードを特定し、前記文書情報格納手段に格納されている文書の中で関連キーワードを含んでいる文書の数を変数とする計算式に基づいて、前記検索条件取得手段が受け取った検索条件と関連キーワードとの間の関連度を算出する関連キーワード計算手段と；前記文書情報格納手段に格納されている文書を対象として、前記関連キーワード計算手段から得られる関連キーワードとその関連度に基づいて、前記検索条件取得手段が受け取った検索条件に関連する関連文書を検索する関連文書検索手段と；前記関連キーワード計算手段から得られる関連キーワードとその関連度に基づいて、前記関連文書の内容から関連箇所を抽出する関連箇所抽出手段と；前記関連箇所抽出手段から得られる関連箇所を出力する関連箇所出力手段とを設けるようにしている。

【0020】この構成においても、関連キーワードとそ

の関連度に基づいて文書を取り出すことができる。関連度は、少なくとも、全文書の中で関連キーワードが出現する割合に基づく。例えば、その割合が小さい場合には情報量としては多くなるので高い関連度を付与する。もちろん、関連する文書における出現割合を加味して関連度を算出してもよい。

【0021】関連キーワードの特定は例えばつぎのように行なうことができる。すなわち、前記検索条件取得手段が、前記文書情報格納手段に格納されている一つあるいは複数の文書を入力として受け取り、前記関連キーワード計算手段が、前記検索条件取得手段が受け取った文書に含まれることばを関連キーワードとし、前記検索条件取得手段が受け取った検索条件と関連キーワードとの間の関連度を、前記文書の中で該関連キーワードを含む文書の数と、前記文書情報格納手段に格納されている文書の中の該関連キーワードを含む文書の数とを変数とする計算式に基づいて算出するようにしてよい。

【0022】また、前記検索条件取得手段が、自然文を入力として受け取り、前記関連キーワード計算手段が、前記文書情報格納手段に格納されているキーワードの中から、前記検索条件取得手段が受け取った自然文に含まれることばを関連キーワードとするようにしてよい。その他種々の手法で関連キーワードを特定することができる。シソーラス辞書を用いて入力キーワードに関連することばを関連キーワードとしてもよい。

【0023】また、前記関連文書検索手段から検索された関連文書を、前記関連箇所出力手段から出力される関連箇所と対応づけて出力する文書出力手段を設けるようにしてもよく、前記関連箇所出力手段および前記文書出力手段が、それらの出力に含まれる、前記関連キーワード計算手段から得られる関連キーワードの出力の形態を、該関連キーワードの関連度に応じて変えるようにしてよい。

【0024】また、本発明によれば、文書検索装置に：複数の文書を、各文書から抽出されたキーワードと対応づけて格納する文書情報格納手段と；検索条件を受け取る検索条件取得手段と；前記文書情報格納手段に格納されている文書を対象として、前記検索条件取得手段が受け取った検索条件に適合する適合文書を検索する適合文書検索手段と；前記適合文書検索手段で検索された適合文書に対応して前記文書情報格納手段に格納されているキーワードを関連キーワードとして求め、前記関連キーワードの各々について前記適合文書の中で該キーワードを含んでいる文書の数と、前記文書情報格納手段に格納されている文書の中で該キーワードを含んでいる文書の数とを変数とする計算式に基づいて、前記検索条件取得手段が受け取った検索条件と該キーワードとの間の関連度を算出し、関連キーワードとその関連度を求める関連キーワード計算手段と；前記関連キーワード計算手段から得られる関連キーワードとその関連度に基づいて、前

(6)

特開2001-84255

記文書情報格納手段に格納されている文書を対象として、前記検索条件に関連する関連文書を検索する関連文書検索手段とを設けるようにしている。

【0025】この構成においては、検索条件に適合しなくとも、関速度が高ければ検索することが可能になる。

【0026】また、本発明によれば、文書処理装置に複数の文章からなる入力文書の各文章を所定の評価基準によりスコアを割り当てる手段と；上記スコアに基づいて順次に文章を抽出する手段と；抽出した文章の累積量が所定の分量基準を上回ったときに抽出を打ち切る手段と；抽出された文章を出力文書にまとめて出力する手段とを設けるようにしている。

【0027】この構成においては、所定の分量基準にしたがって所望の文書部分を抽出することができる。

【0028】この構成において、上記分量基準は入力文書の量の固定割合としてもよい。前記打ち切る手段は、分量基準を上回った時点の文章まで抽出して出力文書に含ませるようにしてもよい。前記出力文書の文章は入力文書における各文章の出現順序で配置されるようにしてもよい。また、前記出力文書の文章は前記評価基準のスコアに基づいて配置されるようにしてもよい。さらに、前記打ち切る手段は、分量基準を上回る直前の文字まで抽出して出力文書に含ませるようにしてもよい。

【0029】また、本発明は方法の発明として実現することもでき、また少なくともその一部をコンピュータプログラム製品（記録媒体）として実現することもできる。

【0030】

【発明の実施の形態】以下、本発明の実施例について説明する。以下の実施例は検索対象の文書として、日本語で記述された文書を想定しているが、その他英語や中国語など、形態素解析を施すことが可能な各種言語で記述された文書に適用することができる。

【0031】【実施例1】まず、本発明の実施例1について説明する。この実施例1は入力した検索文に合致する文書を検索し、文書の各々の関連箇所を抽出するものである。

【0032】図1は実施例1の文書検索装置の構成を全体として示しており、この図において、文書検索装置は、文書情報格納手段1、検索条件取得手段2、適合文書検索手段3、関連キーワード計算手段4、関連箇所抽出手段5、関連箇所出力手段6を含んで構成されている。なお、この実施例の文書検索装置は、コンピュータシステム100（破線で示す）上で実行されるソフトウェアとして実現されている。このソフトウェアは、記録媒体101等を用いてコンピュータシステムにインストールできる。

【0033】文書情報格納手段1は、図5に示すように、単語索引、文書索引等の各テーブル、および文書群を記録するものである。なお、検索対象の文書とは1ま

たは複数の文章からなる検索対象単位であり、1つの処理単位と管理されているものである。文章には1つの単語からなるものも含まれる。

【0034】検索対象の文書には文書IDが与えられており、文書情報格納手段1には、文書IDに対応して文書ファイルが記憶され、さらに、その文書から抽出されたキーワードのリスト（単語索引）と、キーワードに対応してそのキーワードを含む文書の文書IDのリスト（文書索引）とが格納される。キーワードは、検索対象の文書を形態素解析した結果得られた名詞や動詞などの主要な品詞の単語である。

【0035】検索条件取得手段2は、ユーザから検索条件としてキーワードをAND、OR、NOTの論理演算子で結んだキーワードの論理演算式を受け取るものである。

【0036】適合文書検索手段3は、検索条件取得手段2から入力された各キーワードに対応する文書IDのリストを文書情報格納手段1の単語索引から求め、指定の論理演算を施し、適合する文書の文書IDを求める。この文書を適合文書と呼ぶことにする。

【0037】関連キーワード計算手段4は、適合文書検索手段3で検索された適合文書から抽出されたキーワードを関連キーワードとして、各々についてその関速度を計算する。すなわち、関連キーワード計算手段4は、適合文書の各々について、文書情報格納手段1の文書索引を表引きし、それぞれのキーワードを抽出し、これらを関連キーワードとする。そして関連キーワードの関速度は、適合文書の中で、その関連キーワードを少なくとも一つ含む文書の数 α と、文書情報格納手段1に格納されているすべての文書の中で、その関連キーワードを少なくとも一つ含む文書の数 β とを変数とする計算式 $R_w(\alpha, \beta)$ に基づいて算出される。そしてこの計算式 $R_w(\alpha, \beta)$ は、前者の文書数 α の2乗を分子に、後者の文書数 β を分母とする分数で表される。すなわち関連キーワードの関速度 $R_w(\alpha, \beta)$ は、次の式で表わされる。

【0038】

【数1】 $R_w(\alpha, \beta) = \alpha^2 / \beta$

関連箇所抽出手段5は、関連キーワード計算手段4で求まった関連キーワードとその関速度に基づいて、適合文書の各文章について検索条件との間の関速度を計算し、関速度の大きい文章から順に所定の長さ以上になるまで、文書に現れる順番を保ったまま文を抽出していく。文章の関速度は、その文章に含まれている関連キーワードの関速度の合計値である。そして、抽出する文章の最低限の長さは、元の適合文書のテキスト量に対する比率で表わされる固定値である。この値を要約率とよぶことにする。この例では、抽出文の長さの合計が要約率に対応する長さを超えるまで文章を抽出する。もちろん、抽出文の長さの合計が要約率に対応する長さ以内に収まる

(7)

特開2001-84255

ように文章を抽出するようにしてもよい。この例では要約率を10%とした。

【0039】関連箇所抽出手段5は図6に示すように例えばスコア算出手段10、ソート手段11、文章抽出手段12、抽出打ち切り手段13、出力順序付け手段14を含んで構成されている。スコア算出手段10は、文章ごとに関連キーワードの関連度を累積してスコアを算出する。ソート手段11は文章をスコア順にソートする。文章抽出手段12はスコアの高い順に文章を抽出する。抽出打ち切り手段13は、閾値を上回るスコアの文章がなくなったら文章抽出を打ち切る。また、抽出した文章の分量が所定量を超えた場合には、その時点の文章を抽出した後それ以降の文章の抽出をやめる。出力順序付け手段14は抽出した文章の原文における並び順(文章のID)に基づいて文章を並べる。この並びで文章が関連箇所出力手段6から出力される。

【0040】図1において関連箇所出力手段6は、関連箇所抽出手段5から求まる文章を、検索結果としてユーザに対して表示する。

【0041】具体的な検索例について説明する。ここでは用語集(株式会社自由国民社の現代用語の基礎知識1998年版)の各項目を検索対象として、発電に関する話題を検索する状況を想定する。

【0042】検索条件として、キーワード「発電」を説明文中に含む項目を検索した。完全一致検索のヒット件数は図7に示すように61件であった。

【0043】そして、図7に7番目にリストされた、「太陽電池」の項目(文書)の関連箇所(要約)を抽出し、表示することを考える。「太陽電池」についての関連キーワードおよび関連度(スコア)は図8に示すとおりであった。

【0044】そこで、図9に示す「太陽電池」の項目に含まれる全文を文章ごとに区切り、それぞれのスコアを図10に示すように計算する。すなわち、文章ごとに、関連キーワードを抽出しその関連度を累積して対象文章のスコアとする。図10の例では、最初の文章のスコアが「689652」であり、以降、「0」、「0」、「0」、「569826」、「344826」…となる。

【0045】つぎに関連度の大きい順に文章を抽出していく。この例では9番目と1番目の文章を抽出した時点の要約率10%を越え、文書の抽出を打ち切る。この後、抽出した文章をその出現順に並べて出力する。出力結果は図11に示すとおりとなる。

【0046】図11で示す要約において、関連キーワードとしては、「エネルギー」や「エネルギー利用」といったキーワード(説明上図に矢印で示す)が高い関連度をもっていた。

【0047】検索条件で与えたキーワード「発電」は、図9に示す全文には含まれているが、出力された要約

(図11)には含まれていないことが分かる。「太陽電池」の項目を読むと、「発電」ということばは、太陽電池から取り出される電力が交流であると説明している文に使われているが、太陽電池の動作の説明には使われていない。その代わり動作の説明には、「エネルギーの変換」ということばが使われている。検索意図が、厳密に「発電」するものに対してではなく、広く電力を取り出す方法について知りたいという場合、出力された関連箇所はこの検索意図に合致する部分を含んでいる。これは単に検索条件に含まれている単語を含む文を出力した場合には決して得られない効果である。

【0048】なお、関連度の計算式として、上記の式に限らず、たとえば、

【0049】

【数2】

$$Rw(\alpha, \beta) = \alpha / \beta$$

$$Rw(\alpha, \beta) = \alpha^2 / 10g_2(\beta)$$

など、 α と β を引数とする各種の計算式を適用することができる。要するに、関連文書と関連キーワードとの相関を規定する他の式を用いることができる。

【0050】なお関連度の大きい順に文章を並べて要約としてもよい。

【0051】【実施例2】つぎに、本発明の実施例2について説明する。実施例2は、先に説明した実施例1の関連度に基づいて文書を検索し、さらに検索した文書の要約を同様に関連度に基づいて抽出するものである。図2はこの実施例2の構成を全体として示すものであり、この図において図1と対応する箇所には対応する符合を付して詳細な説明を省略する。

【0052】図2において、文書情報格納手段1、検索条件取得手段2、適合文書検索手段3および関連キーワード計算手段4は、図1の実施例1と同じである。

【0053】関連文書検索手段7は、関連キーワード計算手段4から求まる関連キーワードとその関連度に基づいて、文書情報格納手段1から関連文書を検索する。関連文書は関連キーワードを少なくとも一つ含む文書である。また、関連文書の関連度として、その文書に含まれる関連キーワードの関連度を合計した値を計算する。そして、関連文書はその関連度にしたがって順位付けされ、関連度上位から所定の文書数以下で、所定の関連度以上の値をもつ関連文書が、関連度の高い順に検索される。

【0054】関連箇所抽出手段5は、適合文書の代わりに関連文書から文を抽出する他は、実施例1と同じである。関連箇所出力手段6は、実施例1と同じである。

【0055】この実施例では、単にキーワードを用いて検索するのみでなく、キーワードの関連度も考慮しながら検索を行っているのでより検索の意図に合致した項目を検索することができる。そしてこの項目の全文から検索の意図に合致した要約を抽出することができる。

【0056】[実施例3] つぎに、本発明の実施例3について説明する。実施例3は、先に説明した実施例1においてヒットした項目の要約を出力するのみでなく、さらにその全文も表示できるようにしたものである。図3はこの実施例3の構成を全体として示すものであり、この図において図1と対応する箇所には対応する符合をして詳細な説明を省略する。

【0057】図3において、文書情報格納手段1、検索条件取得手段2、適合文書検索手段3、関連キーワード計算手段4、関連箇所出力手段5は、実施例1と同じである。

【0058】文書出力手段8は、関連箇所出力手段5に対応する適合文書の全文を、文書情報格納手段1から求めて表示し、さらに関連箇所に対応する部分を強調表示する。表示の具体例を図12に示す。

【0059】[実施例4] つぎに本発明の実施例4について説明する。この実施例は実施例3と同様に要約文と全文とを出力できるようにしたものである。基本的な構成は実施例3と同様であり、その構成についてはとくに図示しない。

【0060】この実施例においては、関連箇所出力手段6および文書出力手段8が、特定の値以上の関速度をもつ関連キーワードについて、それらの関連キーワードが表示に含まれている場合には下線をつけて強調表示する。表示の具体例を図12および図13に示す。

【0061】[実施例5] つぎに本発明の実施例5について説明する。この実施例は、文書情報格納手段1に格納されている1または複数の文書を指定し、この1または複数の文書と関連する文書を検索するものである。図4はこの実施例5の構成を全体として示すものである。なお、図5において図1または図2と対応する箇所には対応する符合をして詳細な説明を省略する。

【0062】図5において、ユーザは文書情報格納手段1に格納されている1または複数の文書を指定する。検索条件取得手段2は、指定された1または複数の文書IDを入力として受け取る。文書情報格納手段1および関連箇所出力手段6は、実施例1と同じである。関連キーワード計算手段4は、適合文書の代わりに検索条件取得手段2が受け取った文書を用いる他は実施例1と同じである。関連文書検索手段7および関連箇所抽出手段5は、実施例2と同じである。

【0063】この実施例においては、検索条件取得手段2で指定された文書IDに基づいて関連キーワード計算手段4が文書IDで指定された文書のいずれかに含まれるキーワードを関連キーワードとし、その関速度を計算する。この場合、先に説明した関速度の式Rw(α , β)について、検索条件取得手段2に入力した文書IDで規定される1または複数の文書の中で、その関連キーワードを少なくとも一つ含む文書の数を α とし、文書情報格納手段1に格納されているすべての文書の中で、そ

の関連キーワードを少なくとも一つ含む文書の数 β とを変数とする。

【0064】関連文書検索手段7は、文書情報格納手段1に格納されている文書の各々について、それに含まれるキーワードの関速度を累積し、それら文書を関速度にしたがって順位付けし、関速度上位から所定の文書数以下で、所定の関速度以上の値をもつ文書を、検索結果とする。

【0065】この実施例では、文書を指定しその文書と関連する文書を検索することができる。また、この実施例では、実施例1や実施例2で検索したえた文書のうち1つまたは複数のものを指定し関連する文書を検索することができる。もちろんこの実施例の検索で得た関連文書に基づいて検索を行うことができる。このように事前に行った検索結果をもとに繰り返し検索を行える。この際、検索結果から所望のものを指定するようにすれば、検索を最適化することができる。

【0066】[実施例6] つぎに本発明の実施例6について説明する。この実施例は、実施例5が文書情報格納手段1に格納されている文書を指定して関連ある文書を検索したのに対し、自然文を入力し、この自然文に関連する文書を検索するようにしたものである。

【0067】基本的な構成は実施例5と同様であり、図示しない。

【0068】この実施例では、検索条件取得手段2は、検索要求を表現した自然文を受け取る。検索条件取得手段2は自然文について形態素解析をしてキーワードの切り出しを行う。自然文を構成する文章は1つでもよいし、複数でもよい。文章として完結していなくてもよい。この切り出しには通常のパーザーを用いることができる。関連キーワード計算手段4は、切り出したキーワードについて関速度を計算し、これらキーワードの関速度に基づいて文書を選択し関連文書とする。

【0069】関連文書抽出手段5は、選択した関連文書の各々についてキーワードの関速度に基づいて要約を抽出する。

【0070】この実施例によれば、自然文を用いた検索を行え、しかも自然文は検索者の検索意図を的確に表していると考えられるから、検索窓に適合した文書を検索することができる。

【0071】なお、自然文から関連キーワードを生成する代りに、シソーラス辞書を用い入力した言葉から関連キーワードを生成するようにしてよい。

【0072】

【発明の効果】以上のように本発明によれば、

(1) 検索結果の文書の内容の中から検索条件に関連する箇所が抽出される

(2) 出力される関連箇所は、検索条件として明示的に指定されたキーワードだけから求めるのではなく、検索条件に関連する関連キーワードから求めるために、指定

(9)

特開2001-84255

されたキーワードの近傍以外の箇所であっても検索意図に近い箇所が抽出される

(3) 出力される関連箇所は、関連度をもつ関連キーワードに基づいて抽出されるため、関連箇所の中から常に所定の量のテキストが抽出される

(4) ユーザの興味に関する情報をユーザごとにあらかじめ用意しておく必要はないという効果を実現することができる。

【図面の簡単な説明】

【図1】 本発明の実施例1の構成を示すブロック図である。

【図2】 本発明の実施例2の構成を示すブロック図である。

【図3】 本発明の実施例1の構成を示すブロック図である。請求項3に基づく本発明の構成を示すブロック図

【図4】 本発明の実施例1の構成を示すブロック図である。請求項6に基づく本発明の構成を示すブロック図

【図5】 本発明の実施例1の文書情報格納手段1を説明する図である。

【図6】 本発明の実施例1の関連箇所抽出手段5の構成例を説明するブロック図である。

【図7】 検索でヒットした文書のリストを示す図である。

【図8】 関連キーワードのスコアを説明する図である

る。

【図9】 検索でヒットした1文書「太陽電池」の全文を示す図である。

【図10】 検索でヒットした1文書「太陽電池」に確固のスコアを説明する図である。

【図11】 出力された関連箇所の例を示す図である。

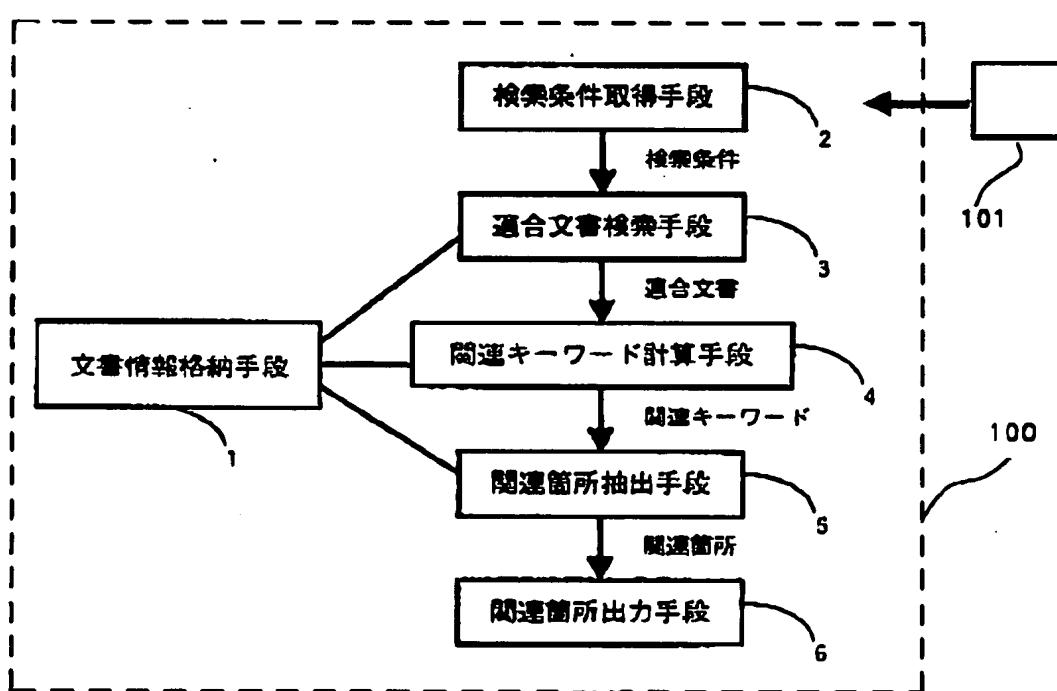
【図12】 関連箇所の表示様式を説明する図である。

【図13】 文書全文の表示様式を説明する図である。

【符号の説明】

- 1 文書情報格納手段
- 2 検索条件取得手段
- 3 適合文書検索手段
- 4 関連キーワード計算手段
- 5 関連箇所抽出手段
- 6 関連箇所出力手段
- 10 スコア算出手段
- 11 ソート手段
- 12 文章抽出手段
- 13 抽出打ち切り手段
- 14 出力順序付け手段
- 100 コンピュータシステム
- 101 記録媒体

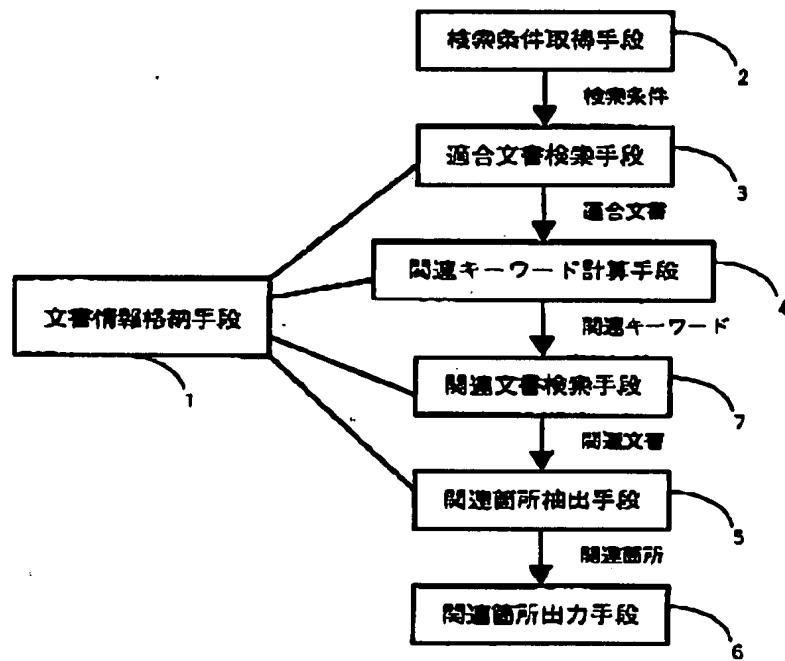
【図1】



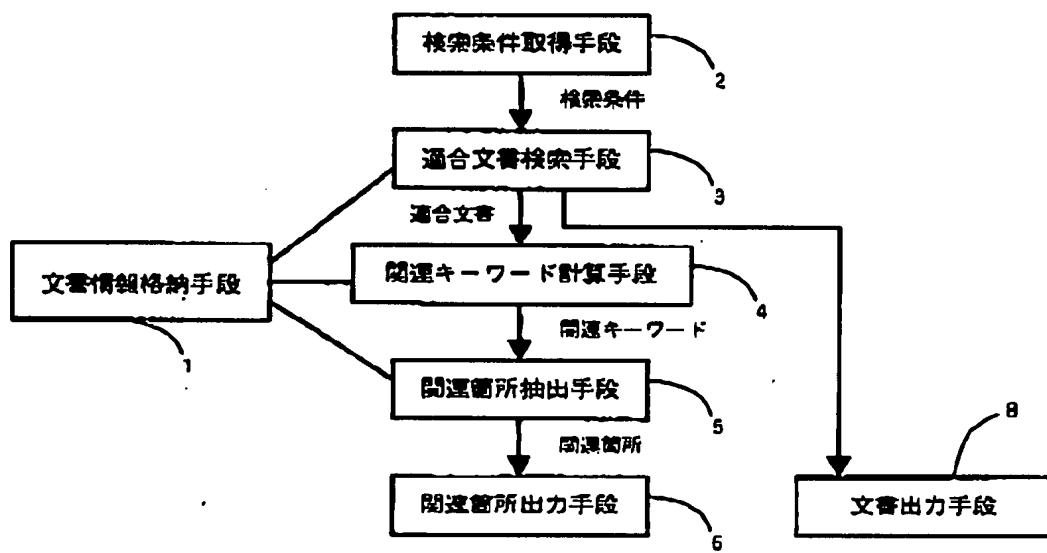
(10)

特開2001-84255

【図2】



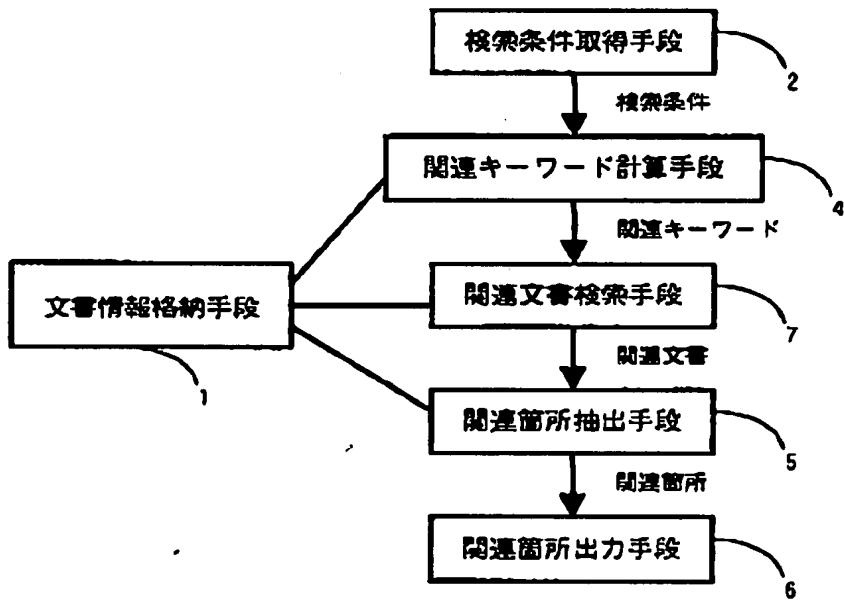
【図3】



(11)

特開2001-84255

【図4】



(12)

特開2001-84255

【図5】

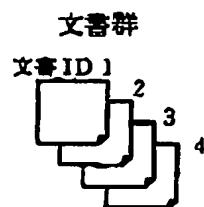
【図5】

単語索引（キーワードと文書IDリスト）

キーワード	対応する文書IDリスト
A	3, 10, ...
AA	4, 6, ...
B	1, 5, ...
BCC	1, 3, ...

文書索引（文書IDとキーワードリスト）

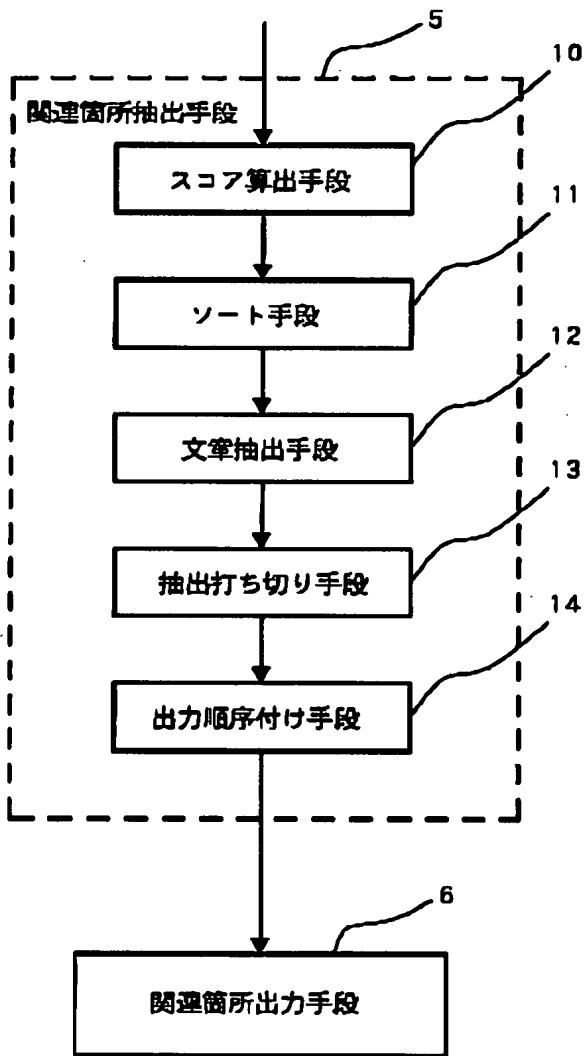
文書ID	キーワードリスト
1	B, BCC, ...
2	C, DD, ...
3	A, BCC, ...
4	AA, C, ...



(13)

特開2001-84255

【図6】



(14)

特開2001-84255

【図7】

完全一致検索結果(61件)のタイトルリスト一覧

- 1.●経済構造改革、同行動計画〔日本経済用語〕
- 2.◆ナホトカ号沈没事故〔旧ソ連用語〕
- 3.★米朝核枠組み協定(US-North Korean Nuclear Framework Agreement)〔平和・軍縮用語〕
- 4.◆スーパー・ゴミ発電〔地方自治用語〕
- 5.◆原発反対闘争〔住民運動用語〕
- 6.◆住宅用太陽光発電システム〔住生活用語〕
- 7.◆太陽電池(solar-cell)〔電子工学用語〕
- 8.◆熱発電素子(thermo electric generating element)〔現代工学用語〕
- 9.◆原子力電源(nuclear power)〔宇宙開発用語〕
- 10.◆(常設)国際宇宙大学(International Space University)〔宇宙開発用語〕
- 11.◆宇宙太陽発電とマイクロ波送電〔宇宙開発用語〕
- 12.◆スペース・コロニー(space colony)〔宇宙開発用語〕
- 13.◆熱水発電(power from hydrothermal vents)〔海洋開発用語〕
- 14.◆海洋エネルギー(ocean energy)〔海洋開発用語〕
- 15.◆波力発電ケーソン(wave power generating caisson)〔海洋開発用語〕
- 16.●深層水の利用(utilization of deep sea water)〔海洋開発用語〕
- 17.★鹿児島県鹿児島地方の地震〔地震・火山用語〕
- 18.◆廃プラスチック(waste plastic)〔化学用語〕
- 19.◆核融合(nuclear fusion)〔物理学用語〕
- 20.○解説の角度〔原子力用語〕
- 21.◆原子力発電所(nuclear power station)〔原子力用語〕
- 22.◆美浜二号機事故(Mihama nuclear power reactor 2 accident)〔原子力用語〕
- 23.◆動力炉(power reactor)〔原子力用語〕
- 24.◆軽水炉(lightwater reactor)〔原子力用語〕
- 25.◆高温ガス炉(high temperature gas cooling reactor)〔原子力用語〕
- 26.◆増殖炉(breeder reactor)〔原子力用語〕
- 27.◆新型転換炉(A.T.R) (advanced thermal converter reactor)〔原子力用語〕

(15)

特開2001-84255

【図8】

関連キーワードとそのスコアの一覧 (スコアの大きい順)

エネルギー用語	544680
エネルギー	635185
エネルギー	635185
発電設備	500000
原子力用語	449438
電気事業	445454
電力会社	426315
原子力発電	422500
燃料電池	367142
ジェネレーション	357142
GENERATION	306250
業者以外	300000
発電所内	300000
事業者	300000
発電功率	300000
火力発電	300000
発電能力	300000
原子力発電所	270000
関西電力	266666
自家発電	225000
電力需要	225000
一次冷却	225000
エネルギー利用	225000
高速増殖	192307
二〇三〇年	180000
プルトニウム	175000

(16)

特開2001-84255

【図9】

シリコン半導体の薄片が示す光電効果を利用したもので、太陽のエネルギーを、直接、電気エネルギーに変換する装置である。太陽電池の素子一個当たりは〇・四～〇・五ボルト、一平方センチ当たり二五ミリアンペアくらいであるが、これをたくさん使って、人工衛星や探査機などの電源として広く利用されている。また身近なものでは太陽電池を組み込んだ電子ウォッチ、屋外時計、電卓、ヒゲソリなども登場してきた。いまのところ太陽エネルギーの約一〇%を電力に変換できる。公害、環境問題、エネルギー問題に関連して、近年、太陽電池実用化の基盤技術にかかる研究が進んでいる。家の屋根に瓦の代わりに並いたとき、どれくらい活用できるかという研究、車に使ったときどれくらい走ることが可能かという研究など、いろいろ工夫されている。なお、太陽電池という用語は厳密には誤りで、エネルギー変換デバイスにすぎない。昔、ベル研究所で最初の発表の時にバッテリーという言葉を使って以来こう言いならわってきた。一九八六年のソ連の Chernobyl 原発事故の教訓から、太陽エネルギー利用の再評価の声が高くなった。太陽電池を、経済性などいろいろな角度から検討する必要が生じている。太陽エネルギーをソーラー・セルで電気に変換する問題はいくつかある。太陽のエネルギーは地球上では希薄なので、どうしても大きな面積でそれを捕えなければならないこと。発電は直流なので、家の中で使う道具によっては、交流に変換することが必要なこと。昼間とり出したエネルギーを蓄えておいて夜使う工夫をして実用化しなければならないこと、などで、これが総合的にまだ高価につく理由である。

項目「太陽電池」の全文

(株式会社自由国民社、現代用語の基礎知識1998年版より引用)

(17)

特開2001-84255

【図10】

「太陽電池」の本文における各文のスコア一覧

689652	シリコン半導体の薄片が示す光電効果を利用したもので、太陽のエネルギーを、直接、電気エネルギーに変換する装置である。
0	太陽電池の素子一個当たりは〇・四~〇・五ボルト、一平方センチ当たり二五ミリアンペアぐらいであるが、これをたくさん使って、人工衛星や無人灯台などの電源として広く利用されている。
0	また身近なものでは太陽電池を組み込んだ電子ウォッチ、屋外時計、電卓、ヒゲソリなども登場してきた。
569826	いまのところ太陽エネルギーの約一〇%を電力に変換できる。
344826	公告、環境問題、エネルギー問題に関するして、近年、太陽電池実用化の基礎技術にかかる研究が進んでいる。
0	家の屋根に瓦の代わりに豆いたとき、どれくらい活用できるかという研究、車に使ったときどれくらい走ることが可能かという研究など、いろいろ工夫されている。
344826	なお、太陽電池という用語は厳密には誤りで、エネルギー変換デバイスにすぎない。
0	昔、ペル研究所では最初の発表の時にバッテリーという言葉を使って以来こう言いならわされてきた。
851076	一九八六年のソ連のチェルノブイリ原発事故の教訓から、太陽エネルギー利用の再評価の声が高くなつた。
0	太陽電池を、経済性などいろいろな角度から検討する必要が生じている。
569826	太陽エネルギーをソーラー・セルで電気に変換する問題はいくつかある。
344826	太陽のエネルギーは地球上では希薄なので、どうしても大きな面積でそれを捕入なければならないこと。
0	発電は直流なので、家の中で使う道具によっては、交流に変換することが必要なこと。
344826	昼間とり出したエネルギーを蓄えておいて夜使う工夫をして実用化しなければならないこと、などで、これが総合的にまだ高価につく理由である。

(18)

特開2001-84255

【図11】

◆太陽電池(solar-cell)（電子工学用語）

【要約(10.2)】シリコン半導体の薄片が示す光電効果を利用したもので、太陽のエネルギーを、直接、電気エネルギーに変換する装置である。一九八六年のソ連のチェルノブイリ原発事故の教訓から、太陽エネルギー利用の再評価の声が高くなった。

項目「太陽電池」の関連箇所
(株式会社自由国民社、現代用語の基礎知識1998年版より引用)

(19)

特開2001-84255

【図12】

◆太陽電池(solar-cell) (電子工学用語)

【要約(10%)】 シリコン半導体の薄片が示す光電効果を利用したもので、太陽のエネルギーを、直接、電気エネルギーに変換する装置である。一九八六年のソ連のチェルノブイリ原発事故の教訓から、太陽エネルギー利用の再評価の声が高くなつた。

項目「太陽電池」の関連箇所
(株式会社自由国民社、現代用語の基礎知識1998年版より引用)

【図13】

シリコン半導体の薄片が示す光電効果を利用したもので、太陽のエネルギーを、直接、電気エネルギーに変換する装置である。太陽電池の粒子一個当たりは〇・四～〇・五ボルト、一平方センチ当たり二五ミリアンペアくらいであるが、これをたくさん使って、人工衛星や無人灯台などの電源として広く利用されている。また身近なものでは太陽電池を組み込んだ電子ウォッチ、屋外時計、電卓、ヒゲソリなども登場してきた。いまのところ太陽エネルギーの約一〇%を電力に変換できる。公害、環境問題、エネルギー問題に関する、近年、太陽電池実用化の基礎技術にかかわる研究が進んでいる。家の屋根に瓦の代わりに並いたとき、どれくらい活用できるかという研究、車に使ったときどれくらい走ることが可能かという研究など、いろいろ工夫されている。なお、太陽電池という用語は厳密には誤りで、エネルギー変換デバイスにすぎない。昔、ベル研究所で最初の発表の時にバッテリーという言葉を使って以来こう言いならわってきた。一九八六年のソ連の Chernobyl 原発事故の教訓から、太陽エネルギー利用の再評価の声が高くなった。太陽電池を、経済性などいろいろな角度から検討する必要が生じている。太陽エネルギーをソーラー・セルで電気に変換する問題はいくつかある。太陽のエネルギーは地球上では希薄なので、どうしても大きな面積でそれを捕えなければならないこと。発電は直流なので、家の中で使う道具によっては、交流に変換することが必要なこと。民間とり出したエネルギーを譲えておいて夜使う工夫をして実用化しなければならないこと、などで、これが総合的にまだ高価につく理由である。

項目「太陽電池」の全文

(株式会社自由国民社、現代用語の基礎知識1998年版より引用)

JP 2001-84255 A5 2004.10.28

【公報種別】特許法第17条の2の規定による補正の掲載

【部門区分】第6部門第3区分

【発行日】平成16年10月28日(2004.10.28)

【公開番号】特開2001-84255(P2001-84255A)

【公開日】平成13年3月30日(2001.3.30)

【出願番号】特願平11-257665

【国際特許分類第7版】

G 06 F 17/30

【F1】

G 06 F 15/403 3 8 0 D

G 06 F 15/40 3 7 0 A

G 06 F 15/401 3 2 0 A

【手続補正書】

【提出日】平成15年10月24日(2003.10.24)

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】請求項16

【補正方法】変更

【補正の内容】

【請求項16】

複数の文書を、各文書から抽出されたキーワードと対応づけて格納する文書情報格納ステップと、

検索条件を受け取る検索条件取得ステップと、

前記文書情報格納手段に格納されている文書を対象として前記検索条件を利用した検索を実行する文書検索ステップと、

前記検索により抽出された文書に対応して前記文書情報格納手段に格納されているキーワードを関連キーワードとして求め、前記関連キーワードの各々について前記適合文書の中で該キーワードを含んでいる文書の数と、前記文書情報格納手段に格納されている文書の中で該キーワードを含んでいる文書の数との少なくとも一方を変数とする計算式に基づいて、前記検索条件取得手段が受け取った検索条件と該キーワードとの間の関連度を算出する関連キーワード計算ステップと、

前記関連キーワード計算手段から得られる関連キーワードとその関連度に基づいて、前記抽出された文書の内容から関連箇所を抽出する関連箇所抽出ステップと、

前記関連箇所抽出手段から得られる関連箇所を出力する関連箇所出力ステップとを有することを特徴とする文書検索方法。